A Point-Based Algorithm for Multiple 3D Surface Alignment of Drug-Sized Molecules

Dissertation

am Fachbereich Mathematik und Informatik
der Freien Universität Berlin

zur Erlangung des akademischen Grades
des Doktors der Naturwissenschaften (Dr. rer. nat.)
vorgelegt von

Diplom-Informatiker
Daniel Baum

Berlin, März 2007

1. Gutachter: Prof. Dr. Dr. h.c. Peter Deuflhard
Fachbereich Mathematik und Informatik
Freie Universität Berlin

2. Gutachter: Prof. Dr. Joachim Selbig
Institut für Biochemie und Biologie
Universität Potsdam

Tag der Disputation: 11.07.2007

Abstract

One crucial step in *virtual drug design* is the identification of new *lead structures* with respect to a pharmacological target molecule. The search for new lead structures is often done with the help of a *pharmacophore*, which carries the essential structural as well as physico-chemical properties that a molecule needs to have in order to bind to the target molecule. In the absence of the target molecule, such a pharmacophore can be established by comparison of a set of active compounds. In order to identify their common features, a *multiple alignment* of all or most of the active compounds is necessary. Moreover, since the "outer shape" of the molecules plays a major role in the interaction between drug and target, an alignment algorithm aiming at the identification of common binding properties needs to consider the molecule's "outer shape", which can be approximated by the *solvent excluded surface.*

In this thesis, we present a new approach to molecular surface alignment based on a discrete representation of shape as well as physico-chemical properties by points distributed on the solvent excluded surface. We propose a new method to distribute points regularly on a surface w.r.t. a smoothly varying point density given on that surface. Since the point distribution algorithm is not restricted to molecular surfaces, it might also be of interest for other applications. For the computation of pairwise surface alignments, we extend an existing *point matching scheme* to surface points, and we develop an efficient data structure speeding up the computation by a factor of three. Moreover, we present an approach to compute multiple alignments from pairwise alignments, which is able to handle a large number of surface points. All algorithms are evaluated on two sets of molecules: eight *thermolysin inhibitors* and seven *HIV-1 protease inhibitors.* Finally, we compare the results obtained from surface alignment with the results obtained by applying an atom alignment approach.

Zusammenfassung

Die Identifizierung neuer *Leitstrukturen* (lead structures) zur Entwicklung optimierter Wirkstoffe ist ein äußerst wichtiger Schritt in der *virtuellen Wirkstoffentwicklung* (virtual drug design). Die Suche nach neuen Leitstrukturen wird oft mit Hilfe eines *Pharmakophor-Modells* durchgeführt, welches die wichtigsten strukturellen wie auch physiko-chemischen Eigenschaften eines bindenden Moleküls in sich vereint. Ist das Zielmolekül (target) nicht bekannt, kann das Pharmakophor-Modell mit Hilfe des Vergleiches aktiver Moleküle erstellt werden. Hier ist insbesondere die *gleichzeitige Überlagerung* (multiple alignment) aller oder nahezu aller Moleküle notwendig. Da bei der Interaktion zweier Moleküle die "äußere Form" der Moleküle eine besondere Rolle spielt, sollte diese von jedem Überlagerungsalgorithmus, der sich mit der Identifizierung von Bindungseigenschaften befasst, berücksichtigt werden. Dabei kann die "äußere Form" durch eine bestimmte Art von molekularer Oberfläche approximiert werden, die man als *solvent excluded surface* bezeichnet.

In dieser Arbeit stellen wir einen neuen Ansatz zur Überlagerung molekularer Oberflächen dar, der auf einer diskreten Repräsentation sowohl der Form als auch der molekularen Eigenschaften mittels Punkten beruht. Um die Punkte auf der molekularen Oberfläche möglichst regulär entsprechend einer gegebenen Punktdichte zu verteilen, entwickeln wir eine neue Methode. Diese Methode ist nicht auf Moleküloberflächen beschränkt und könnte daher auch für andere Anwendungen von Interesse sein. Basierend auf einem bekannten *Point-Matching* Verfahren entwickeln wir einen Point-Matching Algorithmus für Oberflächenpunkte. Dazu erarbeiten wir u.a. eine effiziente Datenstruktur, die den Algorithmus um einen Faktor von drei beschleunigt. Darüberhinaus stellen wir einen Ansatz vor, der Mehrfachüberlagerungen (multiple alignments) aus paarweisen Überlagerungen berechnet. Die Herausforderung besteht hierbei vor allem in der großen Anzahl von Punkten, die berücksichtigt werden muss. Die vorgestellten Algorithmen werden an zwei Gruppen von Molekülen evaluiert, wobei die erste Gruppe aus acht *Thermolysin Inhibitoren* besteht, die zweite aus sieben *HIV-1 Protease Inhibitoren.* Darüberhinaus vergleichen wir die Ergebnisse der Oberflächenüberlagerung mit denen einer Atommittelpunktüberlagerung.

Acknowledgments

The work described in this thesis has been carried out from 2003 to 2007 at the department of Visualization and Data Analysis at the Zuse Institute Berlin (ZIB). I am indebted to my supervisor Prof. Dr. Dr. h.c. Peter Deuflhard and to the head of my department, Hans-Christian Hege, for their continuous support and encouragement. I would also like to thank them not only for giving me the opportunity to graduate, but also for providing me with all the freedom to follow own ideas.

My work was partially supported by the Berlin Center for Genome Based Bioinformatics (BCB), funded by the Federal Ministry of Education and Research (BMBF). I would like thank Frank Cordes, leader of the BCB Junior Research Group "Conformation databases for virtual screening", for his great interest in my work, his stimulating ideas, and for his always open door.

This work was inspired by Daniela Pelz from the Deparment of Neurobiology, FU Berlin, with her work on the olfactory system of Drosophila melanogaster. Beside making me look at structural molecular alignment, I would like to thank her for a fruitful and very enjoyable cooperation.

I am very grateful for the creative and warm atmosphere at the department of Visualization and Data Analysis, and I would like to thank all members of my department for their valuable support. Johannes Schmidt-Ehrenberg deserves special thanks for his always open ear and mind to discuss all matters, no matter whether these were of scientific nature or not.

Some members of the department of Numerical Analysis and Modelling have also contributed to this work. In particular, I would like to thank Susanna Kube and Marcus Weber for their help with mathematical questions. I also owe many thanks to Holger Meyer and Frank Cordes, who provided me with conformational data of the molecules considered in this thesis.

This work would not have been written without all the help and encouragement I received from my family and friends. I consider myself a very happy person to be part of such a wonderful family. Most of all, however, I want to thank Brigitte for her love and understanding, and my children Judith, Luise, and Simon for many wonderful moments.

Contents